Netflix Siblings

Taylor Blair

1/26/2021

Background

My sister and I have a long history of watching Netflix.

There are some minor outliers in the datasets, namely when we use each others accounts.

If you would like to create your own version of this dataset, or find your Netflix history go to: https://www.netflix.com/YourAccount

From there, select the specific profile ⇒ **Viewing Activity** ⇒ **Download All**

Loading Libraries and Data

```
library(tidyverse)
library(openair)
library(readr)
```

```
my_history <- read_csv("~/Coding/R/sis-and-me-netflix/Data/MyNetflixHistory.csv")
sibling_history <- read_csv("~/Coding/R/sis-and-me-netflix/Data/SisNetflixHistory.csv")

sample_n(my_history, 4)
```

```
## # A tibble: 4 x 2
##   Title                                      Date
##   <chr>                                      <chr>
## 1 The Office (U.S.): Season 6: Secret Santa  8/17/19
## 2 Portlandia: Season 3: The Temp             3/22/20
## 3 The Office (U.S.): Season 5: Business Trip 8/7/19
## 4 Portlandia: Season 3: Nina's Birthday      3/22/20
```

Data Prep

As can be seen above, the initial dataset contains only two variables.

Date fixing

```
my_history$Date <- as.Date(my_history$Date,"%m/%d/%y")
sibling_history$Date <- as.Date(sibling_history$Date,"%m/%d/%y")
```

Movie or TV series

```
my_history$type <- grepl(":", my_history$Title)
my_history$type <- my_history$type %>%
  replace(my_history$type==TRUE, "TV Series") %>%
  replace(my_history$type==FALSE, "Movie")

sibling_history$type <- grepl(":", sibling_history$Title)
sibling_history$type <- sibling_history$type %>%
  replace(sibling_history$type==TRUE, "TV Series") %>%
  replace(sibling_history$type==FALSE, "Movie")
```

Series, Season and episode

```
my_history <- my_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")

sibling_history <- sibling_history %>%
  separate(Title,
           c("Series", "Season", "Episode"),
           ": ")
```

Day of Week

Using

```
day_of_week <- function(x){
  return (weekdays(x))
}

days <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")

my_history$day <- sapply(my_history$Date, day_of_week)
my_history$day <- factor(my_history$day,
                         levels = days)

sibling_history$day <- sapply(sibling_history$Date, day_of_week)
sibling_history$day <- factor(sibling_history$day,
                              levels = days)
```

Outputted Tibble

```
sample_n(my_history, 4)
```

```
## # A tibble: 4 x 6
##   Series                Season   Episode            Date       type     day
##   <chr>                 <chr>    <chr>              <date>     <chr>    <fct>
## 1 The Office (U.S.)     Season 6 Mafia              2019-08-16 TV Seri~ Friday
## 2 Ghost in the Shell    SAC_2045 Season 1           2020-06-17 TV Seri~ Wednesd~
## 3 Big Mouth             Season 1 Pillow Talk        2019-10-16 TV Seri~ Wednesd~
## 4 Orange Is the New Bl~ Season 6 Chocolate Chip No~ 2019-04-26 TV Seri~ Friday
```

#Compare dataset

```
my_history$individual <- "Tay"
sibling_history$individual <- "Sis"

merged <- rbind(my_history, sibling_history)
```

Numerical analysis

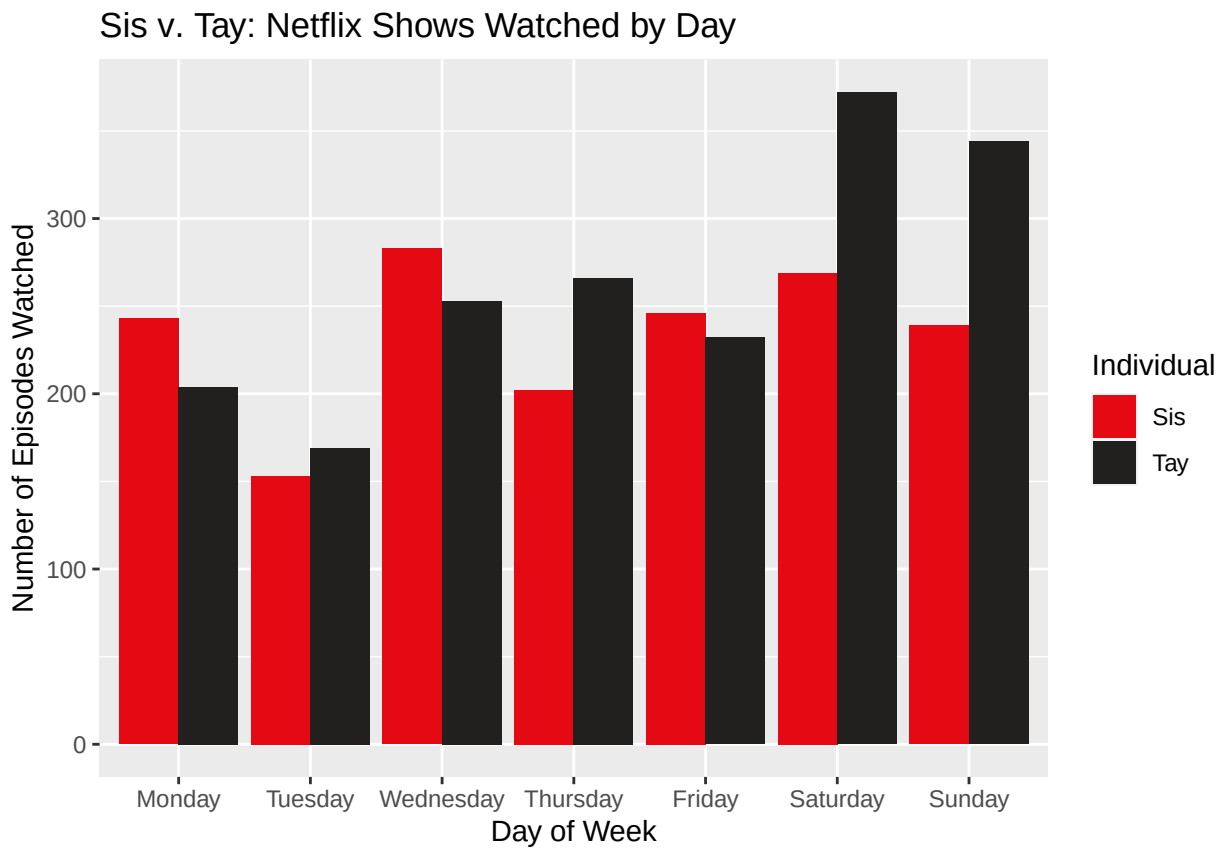
Binge Watching

Who hasn't binge watched Netflix before

Graphical Analysis

Looking at

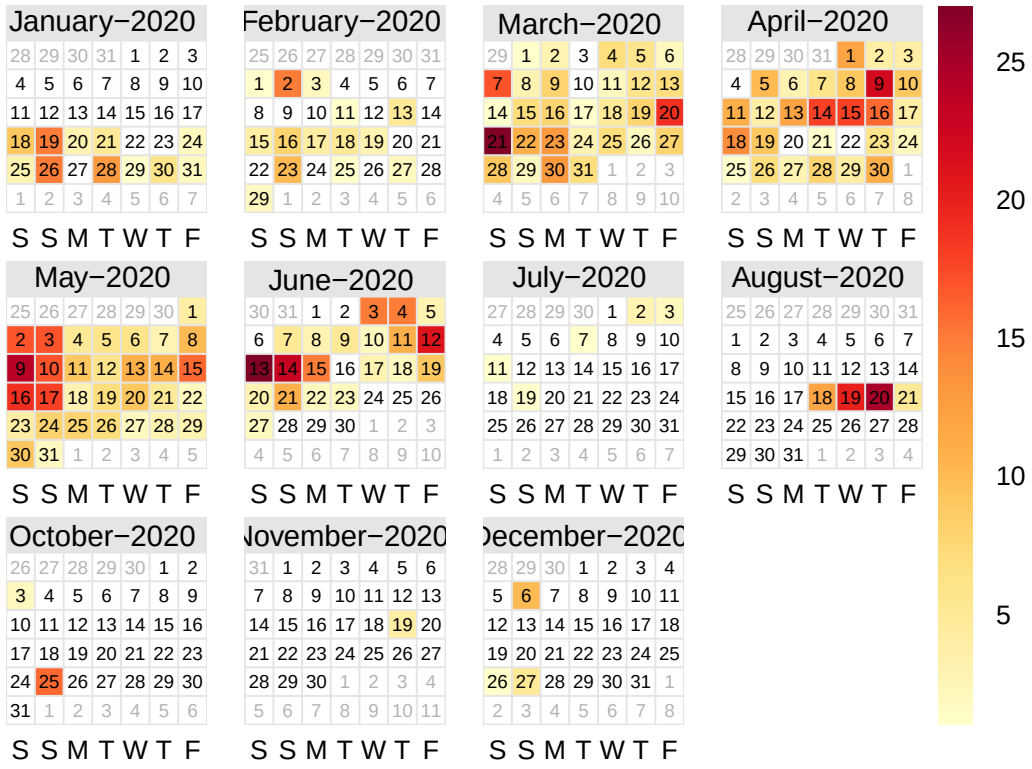
```
ggplot(merged, aes(x = day, fill=individual) ) +
  geom_bar(position = "dodge") +
  labs(title = "Sis v. Tay: Netflix Shows Watched by Day") +
  scale_fill_manual("Individual", values=c("#E50914","#221F1F")) +
  xlab("Day of Week") +
  ylab("Number of Episodes Watched")
```



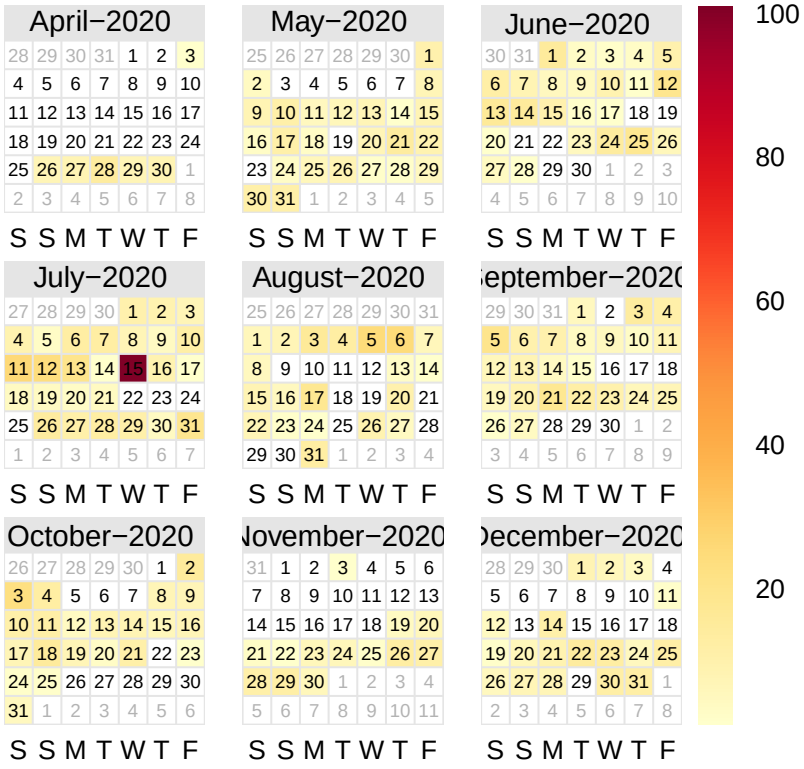
A Calander of Events

2020 was a unique year, with lockdown beginning in February...

```
my_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020)
```




```
sibling_history %>%
  group_by(Date) %>%
  summarise(count = n()) %>%
  rename(date=Date) %>%
  calendarPlot(pollutant = "count",
               year = 2020,
               month = c(1:12))
```


April−2020
28 29 30 31 1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 1
2 3 4 5 6 7 8
S S M T W T F
May−2020
25 26 27 28 29 30 1
2 3 4 5 6 7 8
9 10 11 12 13 14 15
16 17 18 19 20 21 22
23 24 25 26 27 28 29
30 31 1 2 3 4 5
S S M T W T F
June−2020
30 31 1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18 19
20 21 22 23 24 25 26
27 28 29 30 1 2 3
4 5 6 7 8 9 10
S S M T W T F
July−2020
27 28 29 30 1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31
1 2 3 4 5 6 7
S S M T W T F
August−2020
25 26 27 28 29 30 31
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31 1 2 3 4
S S M T W T F
eptember−2020
29 30 31 1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 1 2
3 4 5 6 7 8 9
S S M T W T F
October−2020
26 27 28 29 30 1 2
3 4 5 6 7 8 9
10 11 12 13 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
31 1 2 3 4 5 6
S S M T W T F
November−2020
31 1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 1 2 3 4
5 6 7 8 9 10 11
S S M T W T F
December−2020
28 29 30 1 2 3 4
5 6 7 8 9 10 11
12 13 14 15 16 17 18
19 20 21 22 23 24 25
26 27 28 29 30 31 1
2 3 4 5 6 7 8
S S M T W T F
100
80
60
40
20